FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Delivery Time Slot Management Methods in Online Retail

Armando Silvestre Loureiro Peixoto

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Dr. Sara Sofia Baltazar Martins

Co-supervisor: Dr. Pedro Sanches Amorim

July 23, 2019

Resumo

Com o aumento da adesão às compras online, os retalhistas começaram a praticar serviços de entrega ao domicílio assistidas, o que trouxe desafios em relação às estreitas margens de lucro que esta atividade gera. Este serviço é realizado oferecendo várias janelas temporais para o cliente escolher. Este tipo de serviço prestado torna ainda mais importante otimizar os processos de entrega de modo a manter boas margens de lucro e, ao mesmo tempo, fornecer um bom serviço ao cliente.

Uma vez que são apresentadas janelas temporais aos clientes para escolherem quando fazem uma compra, os retalhistas podem tentar influenciá-los a escolher um determinado horário ao alterar o conjunto de janelas temporais que lhes é oferecido, resultando possivelmente num processo de entrega mais otimizado e económico. A literatura reconhece principalmente quatro tipos de gestão de janelas temporais: janelas temporais fixas, preços fixos, janelas temporais dinâmicas e preços dinâmicos. Este trabalho concentra-se numa estratégia de janelas temporais dinâmicas.

Para conseguirmos ter decisões dinâmicas para janelas temporais eficientes, é necessário fazer-se uma boa estimativa de custo de oportunidade para os clientes que chegam. Calculamos o custo de oportunidade considerando o cronograma de entrega para os clientes já aceites e o futuro cronograma de entrega esperado para os clientes que ainda chegarão no futuro, através de um esquema baseado em sementes. Desta forma, o trabalho de Mackert, 2019 é usado como base para este estudo.

Para resolver este problema, adaptamos estratégias da literatura e apresentamos duas outras que têm como objetivo a satisfação do cliente. Os resultados do nosso trabalho mostram como a aplicação de gestão de janelas temporais pode efetivamente melhorar as margens de lucro de retalhistas online. Além disso, mostramos que, mesmo impondo fatores para a satisfação do cliente, as nossas abordagens levam a resultados muito bons.

Abstract

With the increase of adoption for online shopping, retailers started to practice attended home delivery services, which bring challenges concerning the narrow profitability margins that this activity yield. This service is done by offering multiple time slots for the customer to choose from. This type of service makes it even more important to optimize the delivery processes to maintain good profit-margins while providing a good customer service.

As customers are presented with time slots to choose from when making a purchase, retailers can try to influence customers to choose a certain time slot by changing the set offered to them, possibly resulting in a more optimized and cost-efficient delivery process. The literature acknowledge mostly four types of time slot management: static slotting, static pricing, dynamic slotting and dynamic pricing. This work focus on a dynamic slotting strategy.

In order to have efficient dynamic time slot decisions, it is necessary to make a good opportunity cost estimation for arriving customers. We calculate the opportunity cost by considering the delivery schedule from already accepted customers and the expected future delivery schedule for customers yet to come, by the means of a seed-based scheme. For this purpose, the work of Mackert, 2019 is used as basis for this study.

To solve this problem we adapt strategies from the literature and present two others that target customer satisfaction. The results of our experiment show how the application of time slot management can effectively improve e-grocers profit margins. Additionally, we show that even when enforcing customer satisfaction our approaches leads to very good results.

Acknowledgements

I would like to offer my special thanks to my supervisor, Dr. Sara Martins, for all the dedication, suggestions and specially the constant support and motivation provided, making this stage of my life a lot easier to endure. Also thank you to my co-supervisor, Dr. Pedro Amorim, for all the suggestions and help.

To my parents, sister and family, for helping me to pursue my dreams and encouraging me in all these years. Also to my friends, that always lighten up my days, and to all the teachers and colleagues that accompanied me through this academic journey.

I would also like to leave a special thank you to Inês for always loving, supporting and believing in me.

Armando Peixoto

“Continuous improvement is better than delayed perfection.”

Mark Twain

Contents

List of Figures

List of Tables

Abbreviations and Symbols

AHD	Attended home delivery
GAM	Generalized attraction model
MNL	Multinomial logit
ATC	Area-time slot combination
MILP	Mixed-integer linear program
VRPTW	Vehicle routing problem with time windows

Chapter 1

Introduction

During recent years, e-commerce and the convenience that it carries has allured consumers to opt for this service. Nonetheless, retailers are concerned about the narrow profitability margins that this activity yield, leading them to try to optimize their delivery processes while maintaining a quality service to their clients. In an online business, retailers can have access to plenty information about their clients that otherwise could not, which can be helpful to develop methods and heuristics to improve this service.

For the comfort of the customer, e-grocers normally practices attended home delivery (AHD). This implies that the customer has to be at home to receive their package, hence, the delivery hours are an important feature of the service provided. Unpredictability of demand along with the cost associated to distribution (which in turn rely on the region served and time), affect the cost generated by the AHD services. However, from a customer point of view, the option to choose a particular time slot is convenient, since it minimizes the uncertainty associated with delivery and can choose the most suitable time for him. This way, time slot management is necessary to maintain a good balance between the customer impression towards the service and delivery fulfillment.

To optimize the distribution cost, customers are presented with time slots to choose from when making a purchase, influencing the company's delivery schedule. Thanks to the existence of this choice, retailers can try to influence customers to choose a certain time slot (instead of others that are expected to be more costly to the seller), possibly resulting in a more optimized and cost-efficient delivery process. The selection of these slots can follow a dynamic slotting and/or dynamic pricing strategy. Since the focus of the retailer is to ensure the consumer satisfaction while maintaining a favorable profit margin, it is important to pay attention to customer behavior when defining the time slot management strategy.

There are a lot of studies regarding time slot management, and how we could use it to lead clients to choose more favorable slots. However, the proposed models make several assumptions that are difficult to use in practice, as well as the consumers behavior considered is oversimplified.

1.1 Motivation

Over the years, the online industry have been focused on improving the existent methods of online distribution to increase profitability. While most e-tailors already offer time slots for the customer to choose from, a lot of them use a simple decision tree to select the time slots to offer the customers and are not very efficient. Several approaches to aid this problem have surfaced, such as dynamic slotting of delivery time schedules and dynamic slot pricing, to lead customers to choose more convenient slots.

This project aims to assist the study of this area, where although there is a lot of literature, it is necessary to study different variants, relaxing certain assumptions. The strategies proposed in the literature are usually hard to implement in practice or use pricing strategies that some retailers don't have the ability to perform. As such, there is a need for simpler but effective strategies.

1.2 Objectives

This dissertation will analyze different strategies of time slot management to efficiently assign the more beneficial slots, both from the point of view of the retailer as well as the client, incorporating practical business characteristics.

To portray customer behavior will be applied a model used in the literature, which will support us in predicting the choices made by them through several methodologies that in turn will help us to estimate the opportunity cost of inserting a client in a given delivery route. While a new slotting approach will be proposed for time slot management, state-of-the-art methods will be used to model customer behavior and estimate the opportunity cost.

With this project, we intend to analyze different methods that define strategical rules around time slots that retailers should offer to their clients, with the goal to lead customers to choose more beneficial slots for distribution while ensuring a good service to the customers.

Chapter 2

Literature Review

Literature has been heavily focusing on attended home delivery (AHD) and time slots. This research streams will be discussed in the following sections. Firstly, e-commerce processes are discussed, as well as the AHD problem. Then, a classification of demand management is used to organize the relevant papers.

2.1 E-commerce

In recent years, there has been a substantial growth in e-commerce practices, with an increase adoption both from sellers and customers. Because of this, it is crucial to have an efficient sales strategy, in order to avoid events such as the declaration of bankruptcy by Webvan (Sandoval, 2007).

The way a customer proceeds to make a purchase is usually similar between merchants, with a log in phase which involves customers providing their delivery address, selection of products they want to buy, choosing the payment method and delivery options. The selection of a delivery option, which is the focus of this study, can be carried out before or after the selection of products.

Depending on the sector, sellers may offer different delivery options, such as pick up in store or home delivery, with or without assistance. Home delivery can yet be defined with a delivery time based on deadlines (common for parcel deliveries), such as delivery in 48 hours after checkout, or time slots selection, in which the products have to be delivered at the time pretended by the customer. Some practical examples are Amazon Marketplace that uses a distribution system that doesn't require for the client to be at home, while e-grocers follow an AHD approach.

Logistically, companies also have to make decisions regarding the distribution planning. Some choose to rent services from others distribution companies while others, like e-grocers for example, usually have their own transportation service. Regardless of this decision, the definition of a time slot management approach brings always challenges concerning the route schedule of the orders placed. As such, it is necessary to accommodate every request in different time slots for delivery, while at the same time keeping in mind future customer requests. Furthermore, it can be important to contemplate various variables present on the roads, such as the extent of the area of operation and the traffic volume in suburban or downtown areas (Ehmke and Campbell, 2014). Because of the low profit margins of e-grocers, and since distribution services are expensive (especially when considering AHD with time slots), it is important to perform an efficient time slot management strategy in order to serve more customers with lower delivery costs.

2.2 Time Slot Management

In order to be effective, the time slots presented by retailers have to be managed accordingly. In the literature, we find different approaches to deal with time slot management. Some researches suggest methods to predict and manage time slots per delivery area, before the customer places the order. Thereby, when new customers arrive, are presented to them the time slots previously selected for their area. In the same way, slots can be presented with different prices or discounts to influence the customer. Although these methods result in considerable savings, they consist in approximations. Therefore, Agatz et al., 2013 suggest dynamic approaches to manage demand in real time, shifting the focus from feasibility to profitability. As Klein et al., 2017a point out, we can divide the demand management in the strategies depicted in Table 2.1.

Table 2.1: Classification of Demand Management (Agatz et al., 2013, Yang et al., 2016)

	Time slot allocation	Time slot pricing
Static (offline, forecast-based)	Differentiated slotting	Differentiated pricing
Dynamic (real-time, order-based)	Dynamic slotting	Dynamic pricing

We can choose between anticipate which slots to present (or charge a fee), calculating the opportunity cost of a slot with the use of historical data (static approach), or calculate the opportunity cost of a slot in real-time, i.e., at the moment of checkout (dynamic approach). The former does not provide the best opportunity cost solutions since it is based solely in forecasts. The latter has the disadvantage that it is calculated at the moment of checkout, which can be computationally demanding since it is necessary to present the results in real time. A combination of both could also be an interesting choice.

The key decision in time slot management is to choose between charging an additional fee for an undesirable time slot (to the grocer point of view) or to hide that slot to the customer. Both have their advantages and disadvantages. On one hand, opting for a dynamic pricing approach can be viewed as unfair among customers (Klein et al., 2017b) and, as such, is suggested to give a small discount over full dynamic pricing (Agatz et al., 2013). On another hand, it is also believed that some customers could feel "bad will" from retailers that opt for dynamic slot allocation (Yang et al., 2016). Therefore, there is a trade-off that we need to keep in mind when trying to improve the time slot management strategy implemented.

2.2.1 Differentiated Slotting

Differentiated slotting comes from simple methods of planning in advance the set of time slots to offer to the customer. This methods are not applied in real-time as customers arrive, but instead take into account historical customer data to predict the best set of time slots to offer when a customer eventually arrives, depending on different aspects, such as the area of origin. Literature in this matter has lately brought new ideas to improve and optimize methods for this strategy.

Agatz et al., 2011 propose an approach to help decide which time slots to offer per zip code, in a way to minimize expected delivery costs. To that end, the authors try to optimize the time slot allocation using a continuous approximation approach, which iteratively improves the estimate of the delivery cost of a time slot schedule for a set of zip codes. An integer programming model is also presented as alternative to model the routing costs of the service. At the end, they produce a fully automated approach capable of producing good solutions in short periods of time. They highlight that while short time slots are convenient for the customer, it is more beneficial in a routing perspective to offer larger time slots.

The work of Hernandez et al., 2014 consider a tactical problem where a time slot combination is selected for each zip code and routes are constructed based on those combinations (for each period of the planning horizon). The availability of the combinations are updated as new orders occur. The authors propose two heuristics. The first one is composed of three phases: (i) a creator phase, where a periodic routing plan is built; (ii) a repair phase, where routes from the same period are merged; and (iii) the final improvement phase, where the routes for each day are optimized. The second heuristic propose solves a periodic vehicle routing problem with the addition of time slots, which was proved useful, by improving upon the first heuristic. Nevertheless, the authors highlight that the first heuristic is still competitive under certain circumstances.

Cleophas and Ehmke, 2014 focus on transport capacity to minimize costs of delivery. They resort to the acceptance mechanism presented previously by Ehmke and Campbell, 2014, to improve the initial route calculated based on the forecast of expected delivery requests, by reserving transport capacity for these. Based on the results acquired, the authors reinforce the importance of an accurate forecast of demand, as well as an efficient vehicle routing planning. Their study present a hybrid strategy between a static and dynamic approach.

2.2.2 Differentiated Pricing

Regarding differentiated pricing, literature is not as ample. However, the work of Klein et al., 2017b introduce the tactical problem of this strategy in AHD. Mainly, their objective is to maximize profits by defining different prices per each zip code/time slot combination (ZTC). They propose an exact model approach, accounting for customer choice behavior and incorporating routing constraints, to anticipate the operational routing costs and differentiate time slot pricing. Additionally, they also propose a model-based approximation using a seed-based scheme, with ZTCs assigned to seeds (zip codes). As benchmark, the authors adapt the model presented by Agatz et al., 2011 related with differentiated slotting. This study proved to be successful in increasing e-grocers profits over the practical pricing approaches usually used.

2.2.3 Dynamic Slotting

Campbell and Savelsbergh, 2005 conduct a study with the objective of maximizing the profits of an online grocery service, being one of the first studies on time slot management. The authors consider a case where customers have a time slot profile and the feasibility of accepting an order is based on insertion heuristics. Their results show that dynamically evaluating the feasibility of a delivery is better than limiting the number of deliveries per slot offered. Additionally, they conclude that to achieve better results a fleet composed by many small vehicles should be used as a way to have a more flexible delivery network.

Ehmke and Campbell, 2014 study the problem of home deliveries in metropolitan areas, with the objective to maximize the number of requests accepted for delivery. They use an enhanced feasibility check approach to verify if the request could be accepted without causing late deliveries due to traffic. The authors conclude that the customer location could impact the number of deliveries (and therefore the profitability) of certain areas. When performing a dynamic feasibility check instead of a simple counting of orders to be accepted per slot, it is possible to maximize the number of accepted requests, but still prone to some lateness in the delivery. By considering time-dependent and stochastic travel times, a balance between the number of requests and service quality is achieved. If a fixed buffer time for each customer was used, the lateness would be eliminated, but the number of requests would also be heavily reduced. Moreover, the authors notice that customers tend to choose slots in the afternoon hours, so it is necessary to incentive customers in choosing other slots to improve profit margins.

Köhler et al., 2018 introduce the concept of flexible time slot management on dynamic slotting through customer acceptance mechanisms. This approach aims to offer as many short time slots as possible to maximize the number of requests accepted. The authors propose four different time slot management strategies: (i) offer as default long time slots for the first customers and then, depending on the service condition, offer short time slots if possible; (ii) start by offering the smaller slots and then change to long time slots when it becomes unfeasible; (iii) take into account the location of new requests and offer short time slots when they are in the vicinity of customers already contained in the current route plans; (iv) consider the relative insertion time and relative length of the time span and only offer short time slots when the insertion time of the request does not impact the route plan and the time span continues acceptable. To manage the feasibility of the time slots, the same insertion heuristics presented by Campbell and Savelsbergh, 2005 is used. Their results show that adapting a flexible time slot management can bring benefits to the retailer but at the same time compromise the customers' motivation into going ahead with the order, as they can expect to have a better time slot option at another occasion.

More recently, Mackert, 2019 propose a mixed-integer linear program (MILP) to approximate the opportunity cost and a non-linear binary program to select the most favorable time slots to offer with the objective of maximizing the expected profit. The customer choice behavior is incorporated in the analysis by using a generalized attraction model (GAM). The selection of time slots offered is based on both marginal and displacement costs, considering customer choice behavior and its impact on vehicle routing. The authors prove that this approach outperforms benchmark approaches. Results show that the lower capacity, the higher the gain in profit when using this approach. The size of the delivery areas should not be to big to sustain approximation accuracy and thus, keep the profit performance stable. They conclude that this approach is not suitable to large-scale businesses. Our study is inspired on this work, replicating the same approach to serve as benchmark and extending with new ideas the methods used.

2.2.4 Dynamic Pricing

Studies in dynamic pricing were introduced by Campbell and Savelsbergh, 2006 as a variant of their previous work in dynamic slotting. The authors present the idea of offering discounts or incentives to influence customer's time slot choice. Insertion heuristics are again used to calculate the feasibility of slots, but with the inclusion of new programming methods to model the incentive schemes, considering that many customers are rather indifferent to small differences in delivery prices. Their goal is to manage the feasible slots in a way that increases the probability of a customer choosing a given slot. The authors also simulate an incentive scheme for the customer to choose wider slots. The results demonstrate that with the use of this strategy it was possible to increase the profitability by reducing delivery costs.

Asdemir et al., 2009 suggest a pricing model based on a Markov decision process. The authors consider a setting where the customers can choose from multiple time slots with different prices. The prices are adjusted with a dynamic mechanism based on time, capacity, and demand characteristics of customers, proving that with their model the overall profitability can be improved with a trade-off of possible lost sales.

Yang et al., 2016 show that the overall profitability of distribution can be improved by using historical data to estimate a multinomial logit (MNL) customer choice model. In this work, feasible slots are selected, and then the delivery cost of each slot is estimated based on both the accepted orders, as well as, future expected orders. Two pools of estimated costs are considered: (i) costs based only on the orders already accepted; (ii) costs of the last ten schedules for the same weekday. This work outperformed previous works and the authors state that for dynamic pricing, it is crucial to forecast the expected demand, leading to better results. Later, Yang and Strauss, 2017 extended this work by proposing a new dynamic programming approach to estimate delivery costs. They decompose the delivery problem into smaller problems, one for each area of activity. In this regard, the authors suggest that could be beneficial to offer low delivery charges in areas with low demand density to stimulate demand by the so-called neighborhood effect.

Klein et al., 2017a proposes a new approximation approach by integrating into the existent literature mixed-integer linear programming. Their work consists in progressively adapting the forecast of expected customers with information about the customers accepted to date. Their results demonstrate that this integration with state-of-the-art approaches improves the average expected profit, with even better results for a tight capacity level. This approach is extended by the work of Koch and Klein, 2017 , which introduce the concept of time slot budgets to capture the value of delivery time. Results of this extension show that profits could be further improved under certain assumptions.

We base this study on the work of Mackert, 2019, using his opportunity cost approximation and time slot management problem strategy. We consider the insertion heuristics of Campbell and Savelsbergh, 2004 to check the feasibility of each time slot to offer the customer and to generate the final routes for the delivery. Additionally, we also adopt the seed-based scheme of Klein et al., 2017a to incorporate the cost of serving future customers on the opportunity cost estimation. We will propose two new time slot management strategies that will be compared to the strategies defined by Mackert, 2019.

Chapter 3

Problem Description

This dissertation is based on the case of a Portuguese retailer that practices AHD. It will study how the resource to time slot management methods can help to improve their profits, while trying to ensure the overall satisfaction of its customers.

Usually, the online transaction process goes as follows: the customer first logs into the website with their information and adds the products they want to buy into the cart. Afterwards, they proceed with the purchase, check out and choose a delivery option: pick up in store or home delivery. If they choose home delivery, a delivery time slot from the ones available has to be selected.

The retailer is in charge of managing the routing schedule of online deliveries, which is planned periodically. The fulfillment of the orders is mostly performed through brick-and mortar stores. More precisely, there is a pre-association between certain stores and delivery zones. For deliveries, the retailer has its own fleet.

Their procedure regarding the time slots presented to the customers is done by offering them all the time slots that are feasible to insert them, based on their location. The retailer current procedure of time slots strategy is to limit the number of orders per time slot based on demand forecasts. Time slots might not be feasible for a customer due to: (i) the capacity limit, where there is a limit number of orders that can be accepted for each time slot in the respective area; (ii) the delivery area, where they offer more or less time slots depending on the demand of each area; (iii) a cut-off time, where certain time slots stop being offered after the corresponding cut-off time. All feasible time slots that satisfy these constraints are presented to the customer.

The objective for this dissertation is to make the process more dynamic. Their delivery service consists in a fixed price policy, so the idea will be to develop a method of dynamic slotting. In Figure 3.1 we can see several time slots presented to the customer for him to choose from. More specifically, we see that the delivery schedule for July 17 is already filled in the majority of time slots, while the delivery schedule for July 18 is full on only two time slots and for July 19 the customer has all time slots available. Presented with this scenario the customer chose the time slot of 7 pm to 9 pm for July 17.

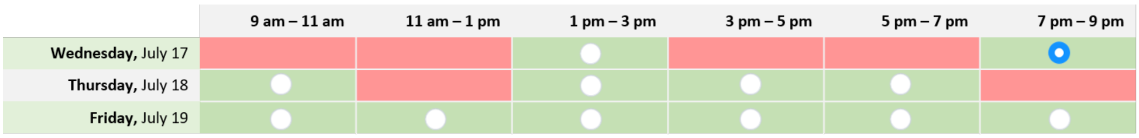

Figure 3.1: Illustrative example of time slots being presented to the customer.

To improve this process and make it more profitable, this study applies the concept of opportunity cost. This value is an estimation of the cost of inserting a costumer in a certain time slot considering the expected customers that could arrive in the future until the end of the booking horizon, i.e., at the time the deliveries of that time slot will be performed, as well as the variations in capacity. This approach allows us to measure the potential profit that certain time slots choices can achieve, and with that we obtain the opportunity cost. Facing the opportunity cost with the revenue of the new request for each time slot, the method decides which time slots to offer the customer. In Figure 3.2 we can see an example of what happens when applying this approach. Although they are feasible (as seen for the same scenario in Figure 3.1), some time slots are not presented to the customer (including the one that he would choose if it was presented to him) and, as such, the customer has to choose another one, changing his option to the time slot 5 pm to 7 pm of July 18.

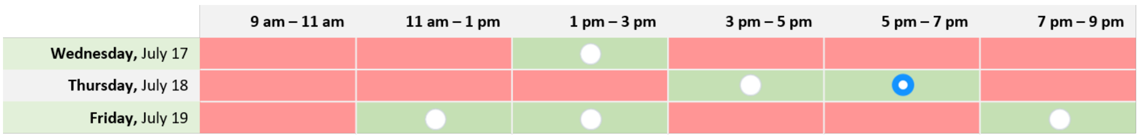

Figure 3.2: Illustrative example of time slots being presented to the customer when considering the opportunity cost.

The study of this problem will be based on the work of Mackert, 2019, that considers customer behavior and dynamic time slot selection, which is the target of our problem. We will use their method for opportunity cost estimation and change their time slot management problem strategy, which is the focus of this study, to further increase e-grocers options. As such, we will present and analyze two new variants to their approaches that aims to improve customer service, and compare to the ones presented by them. On top of that we also define a benchmark approach that tries to resemble the current procedure of the e-grocer under consideration. This leaves us with five approaches to compare:

- Cust-Lim: Benchmark strategy based on the e-grocer current procedure that limits the number of orders per time slot on the delivery schedule.

- No-Opp: The time slot set is determined without considering future customers, so no opportunity cost is considered.

- Opp: The information about expected customers to come is used to estimate the opportunity cost of a customer and shapes the set of time slots that is offered to each request.
- Opp-MinSlot: A variant of Opp that guaranties that a specific minimum number of time slots is offered to the customer whenever possible, i.e. as long as there are enough feasible time slots.
- Opp-MinProb: A variant of Opp that guaranties a specific minimum total probability of customer satisfaction with the time slots offered whenever possible, i.e. as long as there are enough feasible time slots for this purpose.

Chapter 4

Methodology

In this chapter, the heuristic developed to simulate a real case of customers arriving during the booking horizon and placing a request is presented. The model used to decide the best time slots to offer and the model to obtain the opportunity cost of inserting a customer in a route are also portrayed. The methodology is based on Mackert, 2019, with only changes in the time slot offer decision model, depending on the strategy to use.

4.1 Simulation of a Booking Horizon

To analyze each strategy there were performed several customer streams simulations. Each customer stream corresponds to the simulation of a booking horizon divided in T periods. Customers arrive with a rate of λ customers per period with a maximum of one customer per each period t. The logic behind each customer stream can be found in the Algorithm 1 and will serve as foundation for all approaches that will be tested and analyzed in this work.

Algorithm 1 Customer streams heuristic

```
for all cs : customer streams do
    Create historical customers
    for all t : periods do
        if Customer arrives then
            Update seed distances (Section 4.5)
            Thread 1: Calculate opportunity cost (Section 4.3)
            Thread 2: Calculate set of feasible slots (Section 4.4)
            Wait for both function threads to end
            Run time slot management model (Section 4.2)
            Customer selection of time slot (Section 4.6)
        else
            Proceed to next period
        end if
    end for
    Construct final route (Section 4.4)
end for
```

For each customer stream simulation, it is randomly generated a set of customers (historical customers) divided by 12 areas of equal size, on a delivery region of 10 by 10 km. Based on this set, it is defined the probability of a customer arrival be of each area. For each period of the booking horizon, if a customer request arrives it is determined the time slot set to be offered.

Initially the method calculates the opportunity cost of the new request (see Section 4.3) after using a seed-based scheme to estimate the distances of expected future customers (see Section 4.5). Other parameters needed are obtained at the end of the previous period and will be described further ahead. At the same time, it is necessary to look to the already scheduled customers, and verify if it is possible to set up routes for each possible time slot by the means of an insertion heuristic (see Section 4.4). With this it is possible to define the set of time slots that the customer can be feasibly inserted and determine the travel distance required to serve the already accepted customers. Since the calculation of the opportunity cost and definition of the set of feasible time slots take some time to process, and are independent from one another, we can run them in parallel in different threads, sparing some processing time. With all the data collected, we solve the time slot management problem (see Section 4.2) and offer to the customer the most profitable time slots for him to choose from. Afterwards, the customer chooses a time slot based on the underlying customer choice model (see Section 4.6). If a customer does not arrive in that period, the algorithm automatically jumps to the next period. When all the periods are finished, the insertion heuristic used is called again to construct the definitive delivery schedule with all the accepted customers (see Section 4.4).

Figure 4.1 depicts an abstraction of the relational model and data used in this process.

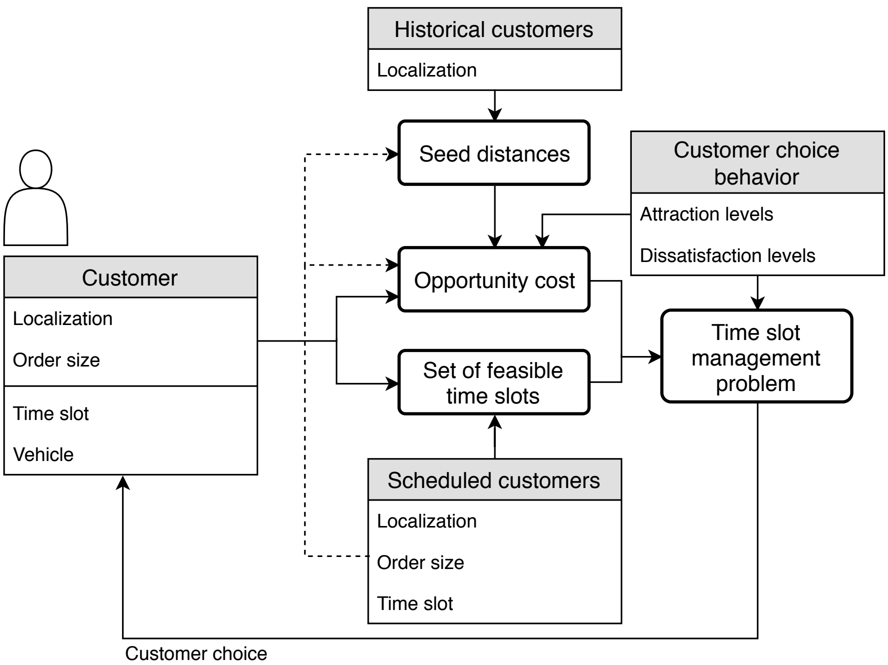

Figure 4.1: Relational model of the used functions and data.

4.2 Time Slot Management Problem

To decide which time slots to offer the customer, we need to account for the feasible time slots $F_a \subseteq \mathscr{S}$ (being $a \in \mathscr{A}$ the area of the customer from the set of service areas $\mathscr{A}$, and $\mathscr{S}$ the set of time slots provided by the service) that the customer can be placed and compare the revenue of that customer with the respective opportunity cost. The customer choice probability is also incorporated to ensure that the decisions take into account the time slots that the customer is more prone to choose.

The objective of this problem is to offer every customer a set of time slots $\mathscr{K}_a \subseteq F_a$, which includes by default the time slot $s = 0$ (no-choice option). For the purpose of this experiment, the calculation of both sets F_a and $\mathscr{K}_a$ are performed only when a customer arrives and, as such, his area is already known. In a real application of this model, these sets can be calculated beforehand with only an estimation of the location of the customer, giving us different solutions to use depending on the customer's area.

This way, the optimal set of time slots to offer $\mathscr{K}_a^*$ is given by

$$\mathscr{K}_a^* = \underset{\mathscr{K}_a}{argmax} \sum_{s \in \mathscr{K}_a} P_{s,a}(\mathscr{K}_a)\left(r + g_s - \left(V_{t+1}(\mathbf{x}) - V_{t+1}(\mathbf{x} + 1_{as})\right)\right). \tag{4.1}$$

The Equation 4.1 aims to find the time slots that maximizes the profit that a customer can bring. It subtracts the opportunity cost to the profit before delivery of the customer, weighted by the probability $P_{s,a}(\mathscr{K}_a)$ of the customer choosing that time slot. The profit before delivery is determined by $(r + g_s)$, in which r represents the current customer order revenue and g_s the fixed delivery fee per time slot. The opportunity cost is defined in the equation by $(V_{t+1}(\mathbf{x}) - V_{t+1}(\mathbf{x} + 1_{as}))$. $V_{t+1}(\mathbf{x})$ represents the value function of the state $\mathbf{x}$, indicating the expected total profit obtainable from period t until the end of the booking horizon if the current customer leaves without placing an order. $V_{t+1}(\mathbf{x} + 1_{as})$ represents the value function of the state $\mathbf{x} + 1_{as}$, expressing the expected total profit obtainable if the current customer from area a is scheduled for the time slot s. How the value functions are calculated is explained in more detail in Section 4.3.

The probability of a customer choosing a time slot ($P_{s,a}(\mathscr{K}_a)$) is defined by the generalized attraction model (GAM) proposed by Gallego et al., 2015. This model allows us to specify different attraction levels for each time slot and area combination (v_{as}), as well as a dissatisfaction level (w_{as}). The purpose of these values are to mimic customer preference for certain time slot, as well as, displeasure if that time slot is not presented. The probabilities are defined as follows.

$$P_{s,a}(\mathscr{K}_a) = \frac{v_{as}}{v_{a0} + \sum_{u \in \mathscr{K}_a} v_{au} + \sum_{u \in \overline{\mathscr{K}_a}} w_{au}} \qquad \forall\, a \in \mathscr{A}, s \in \mathscr{S} \tag{4.2}$$

4.2.1 Solving the Optimization Problem

To solve Equation 4.1, we build the mathematical model (4.3)-(4.5).

$$\max_{\gamma} \frac{\sum_{s\in\mathscr{S}} \left(r + g_s - \left(V_{t+1}\left(\mathbf{x}\right) - V_{t+1}\left(\mathbf{x}+1_{as}\right)\right)\right) v_{as}\gamma_s}{v_{a0} + \sum_{u\in\mathscr{S}} \left(v_{au}\gamma_u + w_{au}\left(1-\gamma_u\right)\right)} \tag{4.3}$$

subject to

$$\gamma_s = 0 \qquad \forall\, s \in \mathscr{S} : s \notin F_a \tag{4.4}$$

$$\gamma_s \in \{0,\ 1\} \qquad \forall\, s \in \mathscr{S} \tag{4.5}$$

The binary decision variable γ_s gives us the solution to the time slot decision problem, and it equals 1 if the time slot s is offered or 0, otherwise.

To the resulting objective function, it is added constraints (4.4) and (4.5). The first one ensures that only feasible time slots can be offered to the customer, while the second defines the values that the decision variable γ_s can take.

A comprehensive notation summary for this mathematical model can be found in Table A.1.

4.2.2 Model Linearization

Since the previous model is non-linear, Mackert, 2019 suggests the following: first, relax constraint (4.5) to $\gamma_s \in [0,\ 1]$, for all $s \in \mathscr{S}$. The author highlights that a local optimum of the problem is a global optimum, so solving the problem with this modification always yields the optimal integer solution. Secondly, it is necessary to linearize the model, and to do so we define the decision variable $\rho := (v_{as} + \sum_{s\in\mathscr{S}}(v_{as}\gamma_s + w_{as}(1-\gamma_s)))^{-1}$ and $\overline{\gamma}_s := \rho\gamma_s$, and modify (4.3)-(4.5) accordingly, resulting in the linear model (4.6)-(4.11).

$$\max_{\rho,\overline{\gamma}} \sum_{s\in\mathscr{S}} \left(r + g_s - \left(V_{t+1}\left(\mathbf{x}\right) - V_{t+1}\left(\mathbf{x}+1_{as}\right)\right)\right) v_{as}\overline{\gamma}_s \tag{4.6}$$

subject to

$$v_{a0}\rho + \sum_{s\in\int} \left(v_{as}\overline{\gamma}_s + w_{as}\left(\rho - \overline{\gamma}_s\right)\right) = 1 \tag{4.7}$$

$$\overline{\gamma}_s = 0 \qquad \forall s \in \mathscr{S} : s \notin F_a \tag{4.8}$$

$$\overline{\gamma}_s \leqslant \rho \qquad \forall s \in \mathscr{S} \tag{4.9}$$

$$\overline{\gamma}_s \geqslant 0 \qquad \forall s \in \mathscr{S} \tag{4.10}$$

$$\rho \geqslant 0 \tag{4.11}$$

With this model, and using an approximation for $V_{t+1}(\mathbf{x})$ (defined in Section 4.3) we can efficiently solve the problem in real-time. Note that the resulting set of time slots is now obtained by dividing ρ to the set $\overline{\gamma}_s$, maintaining the conditions of γ_s.

This is the main decision model for the time slot management decisions, and will have different variations according to the strategy to use. These variations will be discussed in Section 4.7, along with the descriptions of the different strategies.

4.3 Opportunity Cost

To obtain the opportunity cost of inserting a customer in the delivery schedule, we need to calculate the value function for state $\mathbf{x}$ (customer quits the purchase) and each state $\mathbf{x}+1_{as}$ for all $a \in \mathscr{A}$ and $s \in \mathscr{S}$ (customer from area a chooses time slot s). These values mostly measure the impact that a customer has on the delivery schedule, being less restrictive as it approaches the end of the booking horizon. Once again, the model is prepared to calculate these values in advance per each area, giving us several solutions of the value function to use when a customer arrives depending on his location. For the sake of simplification and the nature of the tests made, we calculate each value function only when a customer arrives, since we already have information regarding their location and order size.

Mackert, 2019 gives an approximation for the value function ($\tilde{V}_{t+1}(\mathbf{x})$) in the form of a mixed-integer linear program (MILP), given by the model (4.12)-(4.26). It works as a sort of insertion heuristic that considers both the delivery schedule of already accepted customers, and the expected customers to come. The delivery schedule of already accepted customers defines the parameters d_s^v (travel distance necessary to serve scheduled customers), $\dot{e}_{as}^v$ (order size of scheduled customers) and h_{as}^v (number of scheduled customers). The schedule is determined by solving a vehicle routing problem with time windows (VRPTW) using an insertion heuristic inspired by Campbell and Savelsbergh, 2004 (see Section 4.4). A dynamic seed-based scheme is used to calculate the average travel distances to serve expected future customers (see Section 4.5). The information regarding expected future customers is necessary to determine the opportunity cost, since it is influenced by the impact that the current customer has in the final delivery schedule. This is commonly known as displacement cost, i.e., the insertion of a customer leads to a reduction of capacity levels for future customers that could possibly be more profitable. For this reason, it is approximated for each customer request the final delivery schedule.

The variable $\tilde{s}_{as}$ is set to 1 if we want to calculate the value function of the time slot s for a customer from area a, and to 0 otherwise. The MILP model that approximates the value function is presented next. Table A.2 summarizes the notations used in this model.

$$\widetilde{V}_{t+1}(\mathbf{x}) = \max_{\widetilde{D},\varphi,\delta,\psi,\sigma} \left\{ \sum_{a\in\mathscr{A}} \sum_{s\in\mathscr{S}} \varphi_{as} r - \sum_{s\in\mathscr{S}} \sum_{v\in\mathscr{V}} \widetilde{D}_s^v c \right\} \tag{4.12}$$

The objective function (4.12) aims to maximize the expected total profit from period $t+1$ to the end of the booking horizon T by subtracting the expected delivery cost (expected travel distance $\widetilde{D}_s^v$ times the cost per unit of distance c) to the expected revenue from future customers (future customers φ_{as} time the expected revenue r). The objective function is subject to

$$\frac{v_{a0} + \sum_{s \in \mathscr{S}} w_{as}}{v_{as}} \varphi_{a0} + \sum_{s \in \mathscr{S}} \frac{v_{as} - w_{as}}{v_{as}} \varphi_{as} = \overline{\Phi}_a \qquad \forall a \in \mathscr{A} \tag{4.13}$$

$$\frac{\varphi_{as}}{v_{as}} - \frac{\varphi_{a0}}{v_{a0}} \leqslant 0 \qquad \forall s \in \mathscr{S}, a \in \mathscr{A} \tag{4.14}$$

Constraints (4.13) and (4.14) deal with limits relative to the expected customers and respective time slot choices from period $t+1$ to T, that are dependent of the GAM values. First, constraint (4.13) ensures that the estimated future customers, either leave without placing an order (first term, $s = 0$) or choose a time slot (second term), given by $\overline{\Phi}_a := \lambda \mu_a \Phi$, in which λ equals the arrival rate, μ_a is the probability of customer being of area a, and Φ represents the remaining periods until the end of the booking horizon $T - t$. Constraint (4.14) serves to reflect the fact that the appeal for an alternative time slot offered arises proportionally to its attraction value in relation to the attraction of all other time slot alternatives offered.

$$\sum_{v \in \mathscr{V}} \delta^v = 1 \tag{4.15}$$

$$\varphi_{as} \leqslant \sum_{v \in \mathscr{V}} \psi_{as}^v \qquad \forall s \in \mathscr{S}, a \in \mathscr{A} \tag{4.16}$$

$$\delta^v \widetilde{s}_{as} \leqslant \sigma_{as}^v \qquad \forall s \in \mathscr{S}, a \in \mathscr{A}, v \in \mathscr{V} \tag{4.17}$$

$$\psi_{as}^v \leqslant \sigma_{as}^v M_{as}^v \qquad \forall s \in \mathscr{S}, a \in \mathscr{A}, v \in \mathscr{V} \tag{4.18}$$

Constraints (4.15)-(4.18) manage the routing limitations for each $v \in \mathscr{V}$ (being v a vehicle of the set of vehicles $\mathscr{V}$). Starting with (4.15), it ensures that the current customer request is assigned to exactly one vehicle, given by the binary decision variable δ^v ($\delta^v = 1$ if customer is assigned to vehicle v, 0 otherwise). Next, (4.16) determines that the expected number of customers served between all vehicles for each area-time slot combination (ATC) $(a,\ s)$ (decision variable ψ_{as}^v) has to satisfy all the future requests on the same ATC $(a,\ s)$ (decision variable φ_{as}). The decision variable σ_{as}^v is forced to one if the vehicle v travels to ATC $(a,\ s)$, either to serve the current request (4.17) or the expected customers (4.18). For constraint (4.18) it is also defined the limit M_{as}^v to keep ψ_{as}^v as tight as possible: $M_{as}^v = min\left\{\overline{\Phi}_a \frac{v_{as}}{v_{a0}+v_{as}+\sum_{u \in \mathscr{S} \setminus \{s\}} w_{au}}; \frac{Q-\sum_{b \in \mathscr{A}} \sum_{u \in \mathscr{S}} \dot{e}_{bu}^v}{e}; \frac{l_s-\tau_c \sum_{b \in \mathscr{A}} h_{bs}^v - \tau_d d_s^v}{\tau_c+\tau_d \check{d}_{as}^v}\right\}$, for all $a \in \mathscr{A}, s \in \mathscr{S}, v \in \mathscr{V}$. This value keeps the upper bound of ψ_{as}^v as the minimum out of three possible values: the maximum number of expected customers from area a that can choose time slot s based on the choice probabilities, the maximum number of expected customers that can be served by vehicle v according to the maximum capacity, or the maximum number of expected customers that can be served by vehicle v according to time limitations.

$$d_s^v + \sum_{a \in \mathscr{A}} \widehat{d}_{as}^v \sigma_{as}^v + \sum_{a \in \mathscr{A}} \check{d}_{as}^v (\delta^v \widetilde{s}_{as} + \psi_{as}^v) \leqslant \widetilde{D}_s^v \qquad \forall s \in \mathscr{S}, v \in \mathscr{V} \tag{4.19}$$

Constraint (4.19) further regulates the routing decisions by keeping track of the travel distance of vehicle v in time slot s (decision variable $\widetilde{D}^v_s$). These variable result from the sum of the travel distance for already scheduled customers (d^v_s) with the distance to the current customer's area ($\widehat{d}^v_{as}$), the distance to serve the current customer, as well as all the expected customers ($\check{d}^v_{as}$).

$$\sum_{a\in\mathscr{A}}\sum_{s\in\mathscr{S}}(\dot{e}^v_{as}+\delta^v\widetilde{s}_{as}\dot{e}+\psi^v_{as}e)\leqslant Q \qquad \forall v\in\mathscr{V} \tag{4.20}$$

$$\tau_d\widetilde{D}^v_s+\tau_c\sum_{a\in\mathscr{A}}(h^v_{as}+\delta^v\widetilde{s}_{as}+\psi^v_{as})\leqslant l_s \qquad \forall s\in\mathscr{S},v\in\mathscr{V} \tag{4.21}$$

Both constraints (4.20) and (4.21) contemplates the physical limitations of the problem. Constraint (4.20) ensures that the sum of the order size of the requests already accepted ($\dot{e}^v_{as}$) with the order size of the current request ($\dot{e}$) and expected customers (e), are below the total capacity for each vehicle Q. Constraint (4.21) looks into the limitations of the time slot window. It sums the time used to travel between customers (travel distance $\widetilde{D}^v_s$ times the time needed to travel one unit of distance, τ_d) with the time spent with on-site service (sum of the customers already scheduled h^v_{as}, with the current customer and with the expected number of customers ψ^v_{as}, times the service time needed to serve a customer τ_c) and ensures that it does not exceed the length of the time slot s (l_s).

$$\delta^v\in\{0,1\} \qquad \forall v\in\mathscr{V} \tag{4.22}$$

$$\sigma^v_{as}\in\{0,1\} \qquad \forall s\in\mathscr{S},a\in\mathscr{A},v\in\mathscr{V} \tag{4.23}$$

$$\varphi_{as}\geqslant 0 \qquad \forall s\in\mathscr{S},a\in\mathscr{A} \tag{4.24}$$

$$\psi^v_{as}\geqslant 0 \qquad \forall s\in\mathscr{S},a\in\mathscr{A},v\in\mathscr{V} \tag{4.25}$$

$$\widetilde{D}^v_s\geqslant 0 \qquad \forall s\in\mathscr{S},v\in\mathscr{V} \tag{4.26}$$

The final set of constraints (4.22)-(4.26) defines the decision variables domains. Since the solution depends so much of binary variables, the application is computationally too intense to be solved in real-time. As such, to determine the opportunity cost, the author suggests a small variation to the model by relaxing the binary constraint (4.23). To that end, we substitute (4.23) with two constraints: $\sigma^v_{as}\leqslant 1$ and $\sigma^v_{as}\geqslant 0$, for all $s\in\mathscr{S},a\in\mathscr{A},v\in\mathscr{V}$.

4.4 Insertion Heuristic

The insertion heuristic is used to construct routes for each vehicle $v\in\mathscr{V}$ and time slot $s\in\mathscr{S}$ that the customer can choose from, checking if it can be feasibly inserted. Therefore, we construct a route composed by all previously assigned customers plus the current customer supposing that he chooses time slot $s=1$. After, we do the same supposing that he chooses time slot $s=2$, and so on. In the end, we analyze if the set of routes (for each possible time slot) includes every customer. If at least one customer could not be included, it is not feasible to insert the current customer in the correspondent time slot and, as such, it is not presented in the set of feasible time slots $F_a \in \mathscr{S}$.

Each route begins and ends in the depot, which we consider to be exactly in the center of the delivery region, and no vehicle returns to the depot before servicing all their respective customers. We consider a time window (E_j, L_j) for each customer j, with E_j denoting the earliest time a delivery can take place and L_j the latest time a delivery can take place. We also assume that the delivery period starts at 9 a.m. and ends at 21 p.m. and that the first customer of the first time slot (if there is one) is served at 9 a.m. and last customer of the last time slot (if there is one) can start to be served until 21 a.m. This means that the travel between the depot and first customer is done before the beginning of the delivery period, and the same goes for the travel between the last customer and the depot. Therefore, since each customer is associated a service time τ_c, we consider that as long as the vehicle v serving customer c arrives at that customer before the time slot s window ends, he can be serviced. Figure 4.2 illustrates an example of such a situation.

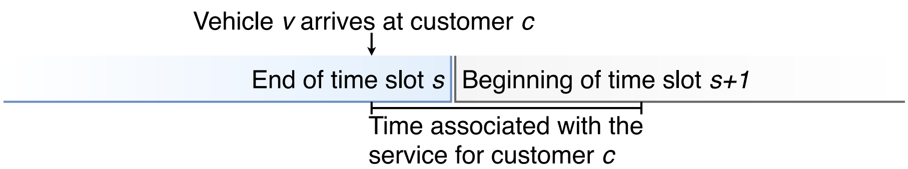

Figure 4.2: Example of a case where service time overlaps two time slot windows.

The insertion heuristic utilized to build these routes is based on the work of Campbell and Savelsbergh, 2004 and presented in more detail in the following sections.

4.4.1 Heuristic

Algorithm 2 presents the main idea of the insertion heuristic. It is run for each time slot, assigning the current customer temporarily to that time slot. It checks for all customers already assigned (i.e. already scheduled for delivery) and the current customer the best feasible and profitable insertion. Each time an insertion is made, the routes have to be updated. If no feasible insertion is found, either for the current customer or any customer of the set of customers already assigned, the algorithm runs again for the next time slot. If for a certain time slot the routes are completed and all customers are scheduled to a route, that time slot can be part of the set of feasible time slots to be offered to the current customer.

Algorithm 2 Insertion heuristic

Input: Set of time slots S; Set of assigned customers N; Set of routes R (one route r for each vehicle); Current customer to be inserted c.

```
for all s : S do
    c.slot ← s
    N ← c
    while N ≠ 0 do
        p∗ ← −∞
        for all j : N do
            for all r : R do
                for all i : r.customers do
                    if FEASIBLE INSERTION(r, i, j) and PROFIT(r, i, j) > p∗ then
                        r∗ ← r
                        i∗ ← i
                        j∗ ← j
                        p∗ ← PROFIT(r, i, j)
                    end if
                end for
            end for
        end for
        if At least one feasible and profitable customer then
            INSERT CUSTOMER(r∗, i∗, j∗)
            UPDATE ROUTE(r∗, j∗)
            N ← N\j∗
        else if There is not a feasible customer in the assigned customers set N then
            Time slot s is not feasible for customer c. Proceed to next s.
        end if
    end while
end for
```

In Algorithm 2 we try to create routes by checking factors such as feasibility and profit of insertion. It is also necessary to update the route every time a customer is inserted. The next group of Algorithms 3-4 explain what happens in those functions.

Algorithm 3 Checking feasibility

```
function FEASIBLE INSERTION(r, i, j)
    e ← max(E_j, e_{i−1} + T_{i−1,j} + service time)
    l ← min(L_j, l_i − T_{j,i} − service time)
    if j.order_size ⩽ available vehicle capacity AND e ⩽ l then
        Return true
    else
        Return false
    end if
end function
```

To check if an insertion is feasible, Algorithm 3 starts by determining the earliest time a delivery can be made to customer j (e_j) and the latest time a delivery can be made to customer j

(l_j) if we insert him between customer i and $i-1$. $T_{a,b}$ is the travel time between customer a and customer b; In this function, the values e and l are temporary and used only in this function. The definitive values of e_j and l_j are determined in the function UPDATE ROUTE, because at this time it is not yet guaranteed that customer j will be placed before customer i.

The insertion is only feasible if the order size of customer j is less or equal to the remaining vehicle capacity available and only if the earliest time a delivery can be made to customer j is less or equal the latest time a delivery can be made to customer j.

$$Profit\,(r,i,j) = -(T_{i-1,j} + T_{j,i} - T_{i-1,i}) \tag{4.27}$$

The profit of an insertion of customer j between customer $i-1$ and i on route r is given by Equation 4.27. The value of the profit is based on the extra travel time introduced by placing customer j between customer i and $i-1$. Since the delivery cost is as higher (and respective profit as lower) as the extra time travel introduced, we use the negative of this value to express a value for profit, to check if it is greater (more profitable) than the last possible insertion.

Algorithm 4 Updating the route

function UPDATE ROUTE($r*$, $j*$)
    for all i : r **do**
        for $k = i-1$ to 0 **do**
            $l_k \leftarrow min(l_k,\ l_{k+1} - T_{k,k+1} - service\ time)$
        end for
        for $k = i+1$ to $r.size$ **do**
            $e_k \leftarrow max(e_k,\ e_{k-1} + T_{k-1,k} + service\ time)$
        end for
    end for
    $available\ vehicle\ capacity \leftarrow available\ vehicle\ capacity - j.order\_size$
end function

Every time a customer is inserted in a delivery route, we need to update the route and respective interactions between customers. For that purpose, in Algorithm 4, we go through each customers i on the route r and update the latest (l_k) time a delivery can be made to the customers k that comes before customer i and the earliest (e_k) time a delivery can be made to the customers k that comes after customer i. Furthermore, we also update the available vehicle capacity to account for the new order.

4.4.2 Preparations for *t + 1* and Final Delivery Schedule

In Section 4.3, it was mentioned that we need the travel distance to serve all scheduled customers until $t-1$ (d_s^v). To get this value we use the delivery route of all scheduled customers and measure the distance between customers. Since in the feasibility check function we build a route for each possible time slot (using the insertion heuristic), we can save the resulting travel distances to use in $t+1$. This way we do not need to run the insertion heuristics (that is considerably heavy in computational effort) more times than necessary. The update of the travel distances is performed at the end of each period after the customer choosing a time slot. If the customer choose any time slot $s \in \mathscr{K}_a$, we update the travel distance d_s^v by the respective value saved from the route where it was inserted. If the customer chooses to leave, the travel distance d_s^v remains the same value as in the last period and it is not updated. This process is repeated for the variables h_{as}^v and $\dot{e}_{as}^v$, since the dynamic nature of the delivery route along the booking horizon will constantly change the set of customers that each vehicle $v \in \mathscr{V}$ will serve.

At the end of the booking horizon $(T+1)$ we run the insertion heuristic one last time to obtain the definitive delivery schedule, i.e., the delivery schedule that will be performed.

4.5 Seed Distances

Since we want to anticipate the delivery schedule and respective delivery cost from $t+1$ until the end of the booking horizon (because of their effect in the opportunity cost), we implement the dynamic seed-based approximation by Klein et al., 2017a. This method results in delivery distances which in turn gives us the delivery cost. They suggest that in each period we dynamically adjust the potential seeds' locations and respective distance approximations. This is required because they will be affected by the location of already accepted customers obtained from the insertion heuristics of Section 4.4.

To calculate the delivery distance of the expected customers from each ATC (a, s), we need to define the respective seed for each vehicle $v \in \mathscr{V}$, which we will use to define the distance necessary for the vehicle v to serve each expected customer in ATC (a, s). This distance is denominated by seed-to-customer distance ($\check{d}_{as}^v$) and is based on the average distance between the seed and the historical customers from area a. Having the distance that a vehicle v has to travel to serve a customer in a ATC (a, s), we have to calculate the distance that the vehicle v needs to travel to arrive at the seed of area a from which it will drive to all expected customers from ATC (a, s). This distance is called to-seed distance ($\hat{d}_{as}^v$).

Each time a customer arrives and chooses a time slot, it influences the delivery schedule and, consequently, the seeds and respective distances. Therefore, we need to recalculate their values in each period t. To approximate these delivery distances, Klein et al., 2017a introduces three different cases to consider for each ATC (a, s).

- Case 1. No customers scheduled in the delivery tour of vehicle v:
  This case applies at the beginning of the booking horizon, when no customers have yet been scheduled in the route for vehicle v and is empty for every time slot $s \in \mathscr{S}$. The seeds for an ATC (a, s) of this case are calculated as the centroid of the historical customers in that ATC (a, s). The seed-to-customer distance $\check{d}_{as}^v$ is obtained by calculating the average distance from this seed to each historical customer for that same area a. The to-seed distance $\hat{d}_{as}^v$ is determined by the distance between the depot and the seed.

- Case 2. There are already customers scheduled in the time slot s in the delivery tour of vehicle v:

In a specific time slot s there is at least one customer scheduled to be served by the vehicle v. For all areas a where there is a customer assigned from case 2, each seed's location is obtained as the centroid of all accepted customer from that ATC $(a,\ s)$. The seed-to-customer distance $\check{d}_{as}^{v}$ is defined as the average distance from this new seed to each historical customer for that same area a. As the vehicle v is already in that ATC $(a,\ s)$ to serve the accepted customer(s), the to-seed distance $\widehat{d}_{as}^{v}$ is equal to zero. If in a certain area a, there are no accepted customers (but they still exist in the same time slot s for other areas a from delivery tour of vehicle v), the seeds for an ATC $(a,\ s)$ are calculated as the centroid of the historical customers in that ATC $(a,\ s)$. The seed-to-customer distance $\check{d}_{as}^{v}$ is obtained by calculating the average distance from the seed to each historical customer for that same area a. Now, the to-seed distance $\widehat{d}_{as}^{v}$ for this particular circumstance is calculated considering an alternative seed that is the centroid of all the historical customer from every area where there are no accepted customers. With this alternative seed, we measure the distance to the seed of each area where there is at least one customer scheduled to be served in time slot s by the vehicle v. With this set of distances, we attribute the shorter one to the to-seed distance variable. This determination is used to reflect that the vehicle v of the existing delivery route should go to this ATC $(a,\ s)$ at minimal cost.

- Case 3. There are already customers scheduled in the delivery tour of vehicle v, but not yet in the time slot s:
  In this case, it is considered all customers that have been accepted in at least one area in any time slot except time slot s. In this situation, the seeds for an ATC $(a,\ s)$ are calculated as the centroid of the historical customers in that ATC $(a,\ s)$ and the seed-to-customer distances $\check{d}_{as}^{v}$ are obtained by calculating the average distance from the seed to each historical customer for that same area a. The to-seed distance $\widehat{d}_{as}^{v}$ is the distance between the seed and the centroid of all customers scheduled in the closest time slots to time slot s. For instance, if we are considering time slot $s = 3$, we first analyze if there are customers in time slot $s = 2$ and $s = 4$, if there is not a single customer in each of these time slots, we look into time slot $s = 1$ and $s = 5$, and so on.

Every distance is equal to the Euclidean distance between both points, multiplied by a factor of 1.5 as a way to approximate a road network (Mackert, 2019).

4.6 Customer Selection of a Time Slot

After the selection of the time slots to offer, the customer is presented with the set of time slots $\mathscr{K}_a$ to choose from. We simulate the customer choice recurring to the GAM. For that, first we calculate the choice probability $P_{s,a}(\mathscr{K}_a)$ of each time slot using the definition 4.2 and the respective no-purchase probability $P_{0,a}(\mathscr{K}_a)$ given by Equation 4.28.

$$P_{0,a}(\mathscr{K}_a) = \frac{v_{a0} + \sum_{u \in \overline{\mathscr{K}_a}} w_{au}}{v_{a0} + \sum_{u \in \mathscr{K}_a} v_{au} + \sum_{u \in \overline{\mathscr{K}_a}} w_{au}} \qquad \forall\ a \in \mathscr{A}, s \in \mathscr{S}. \tag{4.28}$$

Secondly, we randomly select one time slot $s \in \mathscr{K}_a \cup \{0\}$ weighted by the respective probability.

With the time slot s chosen, we proceed to assign the customer to that time slot and update the information impacted by accepting that customer, as referred in Section 4.4.2.

4.7 Time Slot Management Strategies

As previously mentioned, different strategies are considered in this study. Their variations change the resulting set of time slots to offer. The main differences between strategies are the consideration of the opportunity cost (considering the expected future customers) and the service level imposed to the time slots offered.

- Cust-Lim: Does not use any optimization model and only checks the feasibility of inserting a customer, as described in Section 4.4. It defines a maximum number of orders that can be scheduled in each time slot.

- No-Opp: The time slot set is determined without considering future customers. As such, the opportunity cost $((V_{t+1}(\mathbf{x}) - V_{t+1}(\mathbf{x} + 1_{as})))$ used in the objective function (4.6) is replaced by zero in this case, for all time slots, in the problem defined.

- Opp: It is used the value function approximation defined in Section 4.3 to estimate the opportunity cost of a customer, and shape the set of time slots that is offered to him.

- Opp-MinSlot: It is a variant of Opp that guaranties that a specific minimum number of time slots is offered to the customer whenever possible, i.e. as long as there are enough feasible time slots, otherwise it is offered all feasible. For this variant, we incorporate the constraint (4.29) in the model (4.6)-(4.11).

$$\sum_{s \in \mathscr{K}_a} \bar{\gamma}_s \geqslant \left(N\beta + \dot{N}(1-\beta)\right)\rho \tag{4.29}$$

  Equation (4.29) ensures that the total number of time slots $s \in \mathscr{K}_a$ offered to the customer is greater or equal the number N defined by the retailer. If this number is greater than the total number of feasible slots that can be offered to the customer ($\dot{N}$), then this is the value used for minimum number of time slots. We use the binary auxiliary variable β to determine which of the lower bounds is used. Note that, because of the linearization performed to obtain the model(4.6)-(4.11), we multiply the second side of the equation by ρ so that both sides can be effectively comparable with each other.

- Opp-MinProb: It is a variant of Opp that guaranties a specific minimum probability of customer satisfaction with the time slots offered whenever possible, i.e., as long as there are enough feasible time slots for this purpose, otherwise it is offered all feasible. We link the customer satisfaction to the probability of him choosing a time slot $s \in \mathscr{K}_a$ that is obtained with the GAM and is given by (4.2). In fact, since these probabilities use the attraction level of the customer towards a certain time slot s and depends on which time slots are actually presented to him, we can claim that the total probability of him choosing a time slot from the set offered to him gives an approximation of his satisfaction. For this variant we incorporate the constraint (4.30) in the model (4.6)-(4.11).

$$\sum_{s \in \mathscr{K}_a} v_{as}\overline{\gamma}_s \geqslant P\beta + \dot{P}(1-\beta) \tag{4.30}$$

In light of the linearizaton performed in Section 4.2, we know that the probability of a customer choosing time slot s is given by $v_{as}\overline{\gamma}_s$, and as such, the constraint (4.30) ensures that the sum of the probabilities of the time slots there are presented to the customer is greater or equal to either the minimum probability defined by the retailer P or the minimum probability possible given the set of feasible time slots ($\dot{P}$). Once again, we use a binary auxiliary variable β to determine which of the lower bounds is used.

Chapter 5

Results

This chapter presents the results that were achieved by the application of the approaches analyzed, as well as other used as benchmark. First, in Section 5.1 the scenario to which the tests were performed is presented. In the end, in Section 5.2, the results of each approach and respective interpretation are presented.

5.1 Data Description

For our experiment, we try to run a scenario as close as possible to the one used by Mackert, 2019, so most values are directly transcribed of his work.

For each approach we run 50 customer stream simulations, with a booking horizon composed of $T = 700$ periods in each one of them, and a customer arrival rate of $\lambda = 0.814$ per period $t \in \{1, \cdots, T\}$. Each delivery day is divided by 6 time slots with a length of $l_s = 2$ hours and a fixed cost of $g_s = €\,3$ for all $s \in \mathscr{S}$. The grocer has at his disposal a fleet of 5 vehicles with a maximum capacity of 140 totes per vehicle. The customer's order size follows a normal distribution with a mean of 3 totes and a standard deviation of 2. The revenue before delivery is 30% of € 30 (€ 9) per tote. As such, the revenue r used in the time slot decision problem presented in Section 4.2 is the result of the revenue per tote times the order size of the current request. For the opportunity cost model presented in Section 4.3, the revenue r is the result of the revenue per tote times the mean customer's order size. The service time per customer τ_c is equal to 0.2 hours, the time needed to travel one unit of distance τ_d is equal to 0.03 hours per km and the delivery cost c is equal to € 0.3 per km.

The dimension of the delivery region is equal to 10 km by 10 km which is divided in 12 rectangular areas of equal size. At the beginning of each customer stream, 1000 customer are randomly generated and distributed between all areas. This information is after used to give the approximate probability of an arriving customer being from area a, (μ_a). The exact location of these customers is also essential to estimate the distances for the seed-based scheme.

Respecting the customer behavior, attraction levels are determined by Yang et al., 2016 based on real data and are presented in Table 5.1. The attraction level for the no-purchase option v_{a0} for all $a \in \mathscr{A}$ is equal to 1 and the dissatisfaction level is not considered in our problem, hence $w_{as} = 0$ for all $a \in \mathscr{A}, s \in \mathscr{S}$. It is possible to have different attraction levels for each area, but we assume the same values for every area.

Table 5.1: Attraction levels for 6 time slots.

	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$	$s = 6$
$v_{as}\ \forall\ a \in \mathscr{A}$	0.267	0.300	0.188	0.147	0.162	0.179

Lastly, we calculate the profit before delivery as the sum of the revenue of each accepted customer (the average profit per tote times the customer's order size) plus the delivery fee. The delivery cost is calculated by measuring the distance between all customers from each route, including the departure and the arrival at the depot, times the cost per unit of distance. The total profit is obtained by subtracting the delivery cost to the profit before delivery.

This methodology was implemented in C++ using Visual Studio 2017 with IBM CPLEX 12.9 libraries for the mathematical models.

5.2 Simulation Results

By applying the methodology developed to the five time slot management strategies, it was obtained the results presented in Table 5.2. For the Opp-MinSlot strategy we define a minimum limit of 2 time slots to offer to the customer whenever possible, and for Opp-MinProb we ensure a minimum probability of customer satisfaction by the time slots offered of 25% whenever possible. In benchmark strategy Cust-Lim we define a maximum limit of 8 orders per time slot for each route.

Table 5.2: Results for different strategies.

Slotting approach	Average number of scheduled requests	Average number of slots offered	Profit before delivery	Delivery cost	Total profit	Gap*
Cust-Lim	211.50	3.75	6468.21	169.38	6298.84	-5.13%
No-Opp	215.88	3.85	6606.18	167.36	6438.82	-3.02%
Opp	205.10	3.39	6792.18	169.70	6622.48	-0.25%
Opp-MinSlot	208.36	3.41	6753.00	168.33	6584.67	-0.82%
Opp-MinProb	211.16	3.43	6805.32	166.22	6639.10	0.00%

*Relative to best total profit value.

With the results presented in Table 5.2 we can see that relative to the method closer to the currently being used by the e-grocer, Cust-Lim, a simple method of insertion heuristics without limiting the customer per time slot like No-Opp, can improve considerably their profit-margins as well as the number of accepted customers.

In regard to the opportunity cost concept, its application clearly improves even more the e-grocers profit-margins. Looking to Opp, we see that the total profit increases relatively to No-Opp, at the expense of accepting less customers. The reason for the decrease in number of scheduled customers is directly related to average number of time slots offered, that is equally lower. This is explained by the maximum capacity of each route being reached more quickly on the account of the higher average customer's order size on which the application of the opportunity cost take preference.

The Opp-MinSlot approach results in a solution between No-Opp and Opp, which was expected. We see that the number of scheduled requests is somewhat higher compared to Opp. However, since this was caused by the fact that there were more customers with low opportunity cost being offered more time slots (at least 2) than would be with the Opp approach, it does not result in higher profit.

Surprisingly, the application of Opp-MinProb gives the best results overall. This can be explained by the higher number of scheduled customers and respective increment of delivery fees. In fact, if we do not include a delivery fee, the approach Opp gives the best results (a profit without delivery fees of 6007.18 for Opp as opposed to a profit without delivery fees of of 6005.62 for Opp-MinProb), even though it has less scheduled requests. This means that the average order size and respective profit per customer is higher by applying Opp instead Opp-MinProb. The increase of customers scheduled is caused by the fact that not only more time slots are offered to them, but they also are the ones that appeal more to customers. Since this approach tries to offer the best time slots in terms of opportunity cost that satisfies the minimum probability given, it results in a good balance between overall profit and customer satisfaction.

To visualize the evolution of the schedules in the booking horizon, and to better understand the behavior of the different strategies, we selected one customer stream as example and present the evolution of the customers scheduled and the profit before delivery in Figure 5.1 and 5.2, respectively. These analysis only focus on the last four strategies.

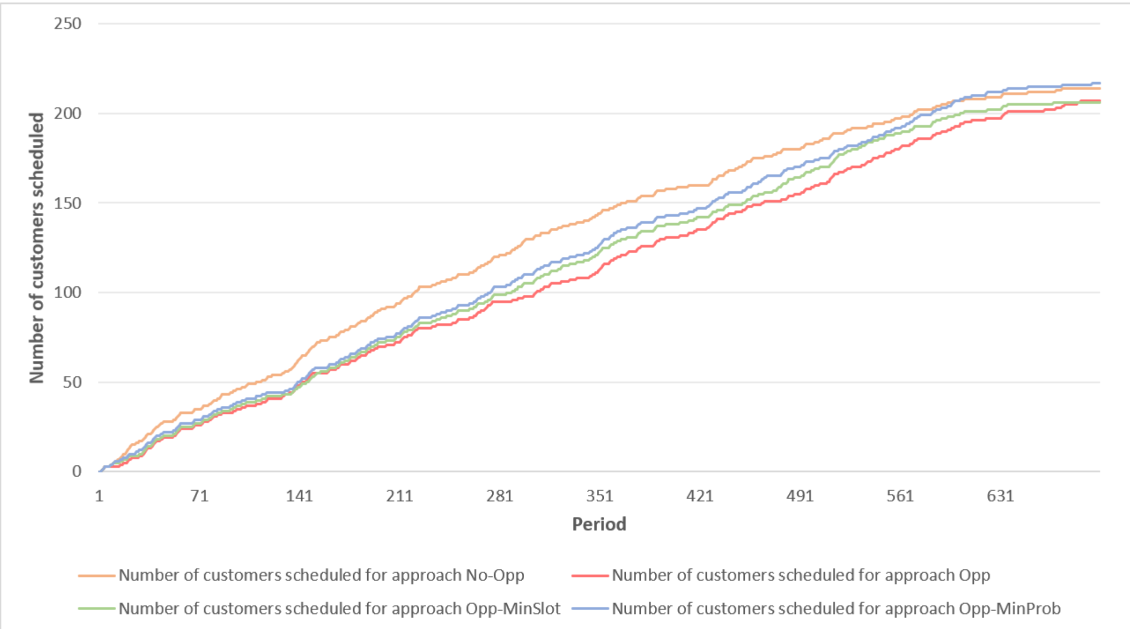

Figure 5.1: Example of customers scheduled evolution for one customer stream.

In Figure 5.1 we see that the number of customers scheduled increases more rapidly with the No-Opp approach during a majority of the periods (more or less until period 491 and 561). A reason for this behavior is that in this approach we do not control the time slots offered to the customers and as such all feasible time slots are offered to him. On the other hand, for the other three strategies that incorporates the opportunity cost, less time slots are offered to the customer even though more feasible time slot are still available, leading to less customers choosing a time slot. However, as we get closer to the end of the booking horizon, in No-Opp we see the number of customers scheduled stagnating somewhat sooner than the other strategies, that still have some capacity left thanks to the more selective offer made during a large part of the booking horizon. These approaches (Opp, Opp-MinSlot and Opp-MinProb), are not as restrictive at the end of the booking horizon, offering mostly all time slots available, as there is a lower expectation of customers to come.

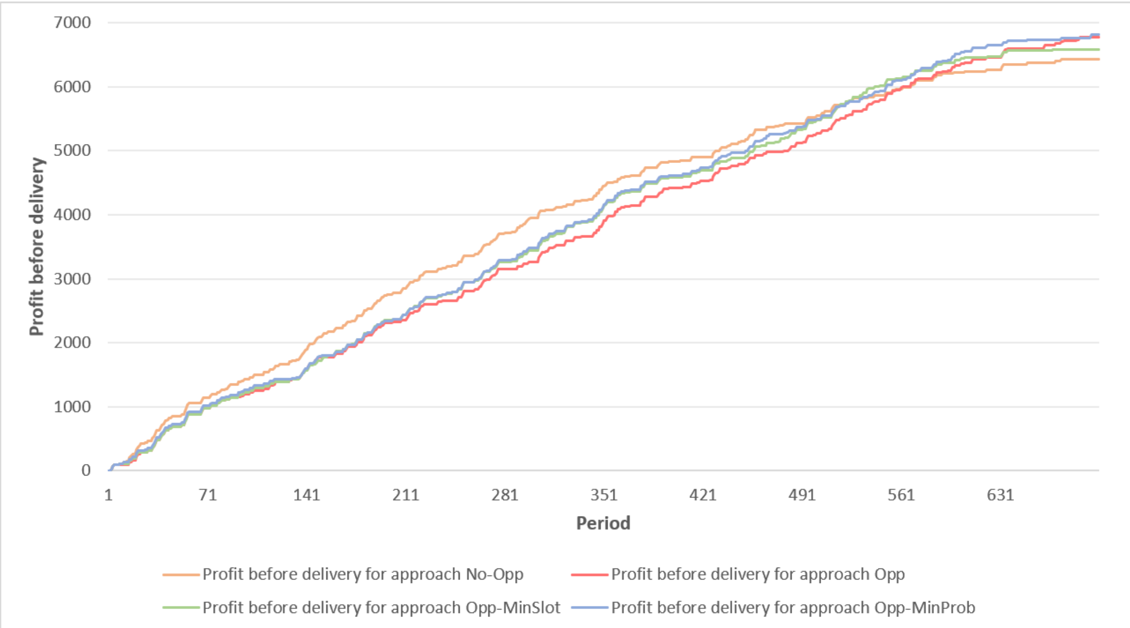

Figure 5.2: Example of profit before delivery evolution for one customer stream.

Looking to the profit evolution, Figure 5.2 shows almost the same behavior than Figure 5.1. This is expected as the profit is associated with the number of accepted customers. Looking between period 1 and 491 of the booking horizon, we see that in fact the profit before delivery obtained by the customers accepted through the No-Opp approach is greater than the one obtained through the other three approaches. However, this difference is smaller than the difference between number of customer observed in Figure 5.1. This is also expected since the mean order size of the customers for Opp, Opp-MinSlot and Opp-MinProb is expected to be higher as a result of the application of the opportunity cost. This way, the difference between approaches regarding profit is not as noticeable as the difference between the number of customers accepted.

In the end of the booking horizon the profit starts to stagnate in all approaches. For Opp,

Opp-MinSlot and Opp-MinProb, on top of less customers being accepted, the selection is less restrictive regarding the request's order size, offering less profitable time slots, as there is a lower expectation of customers to come.

Chapter 6

Conclusions

This work provides an analysis of different strategies to solve the problem of time slot management in online retail. The goal was to maximize the overall profit of an e-grocer by the use of methods that considered customer behavior and also to improve these methods by offering solutions for customer satisfaction.

To solve this problem, the methodology proposed used an opportunity cost model presented by Mackert, 2019 to evaluate how much an insertion of a new customer would benefit the delivery schedule and overall profit. The delivery schedule was determined based on an insertion heuristic adapted from Campbell and Savelsbergh, 2004. Finally, a time slot decision model proposed by Mackert, 2019 was applied, incorporating the opportunity cost and the information about the time slots that the new request could be feasibly inserted in. The results show that the introduction of the opportunity cost concept fairly improves the overall profit. We improved this model to contemplate customer satisfaction, and provide an approach that joins the opportunity cost concept with a minimum probability of customer satisfaction by the time slots offered to him. This approach proved to be successful by attracting more customers and resulting in a good relation between overall profit and customer satisfaction.

This work was selected to be presented in the XX Congress of the *Associação Portuguesa de Investigação Operacional* (APDIO) through the StudIO iniciative. This reinforces the strong potential of this study and in the future we hope to publish a paper with more improvements on this work.

Future directions include a reformulation of the customer choice behavior model to accommodate for new assumptions in the estimation of the opportunity cost, such as the day-dependency of the customer choice to acknowledge that a client can be flexible with the day and have preference only at a specific time.

Appendix A

Notation Summary

A.1 Notation Summary for Model (4.3)-(4.5)

Table A.1: List of sets, input parameters, and decision variables for the time slot offer decision problem.

Sets	
$\mathscr{S} = \{1, \cdots, S\}$	Set of time slots
Parameters	
r	Current customer's profit before delivery
g_s	Delivery fee for all $s \in \mathscr{S}$
v_{as}	Attraction value for time slot $s \in \mathscr{S} \cup \{0\}$ of customers from area $a \in \mathscr{A}$ including the no-purchase alternative
w_{as}	Level of dissatisfaction if time slot $s \in \mathscr{S}$ is not offered to customers from area $a \in \mathscr{A}$
Decision variables	
$\gamma_s \in \{0, 1\}$	$= 1$, if time slot $s \in \mathscr{S}$ is offered to the customer, 0 otherwise

A.2 Notation Summary for Model (4.12)-(4.26)

Table A.2: List of sets, input parameters, and decision variables for the value function.

Sets	
$\mathscr{A} = \{1, \cdots, A\}$	Set of areas
$\mathscr{S} = \{1, \cdots, S\}$	Set of time slots
$\mathscr{V} = \{1, \cdots, V\}$	Set of vehicles
Parameters	
c	Cost per unit of travel distance
d_s^v	Travel distance of accepted and scheduled customers in time slot $s \in \mathscr{S}$ served by vehicle $v \in \mathscr{V}$

$\widehat{d}^v_{as}$	To-seed distance of vehicle $v \in \mathscr{V}$ to travel to ATC (a,s) for all areas $a \in \mathscr{A}$ and time slots $s \in \mathscr{S}$
$\check{d}^v_{as}$	Seed-to-customer distance of vehicle $v \in \mathscr{V}$ to serve an expected customer in ATC (a,s) for all areas $a \in \mathscr{A}$ and time slots $s \in \mathscr{S}$
$\dot{e}^v_{as}$	Order size of scheduled customers in ATC (a,s) for all areas $a \in \mathscr{A}$ and time slots $s \in \mathscr{S}$ served by vehicle $v \in \mathscr{V}$
$\dot{e}$	Order size of current customer request
e	Expected customer's order size
h^v_{as}	Number of scheduled customers in ATC (a,s) for all areas $a \in \mathscr{A}$ and time slots $s \in \mathscr{S}$ served by vehicle $v \in \mathscr{V}$
l_s	Length of time slot $s \in \mathscr{S}$
M^v_{as}	Sufficiently large number for area $a \in \mathscr{A}$, time slot $s \in \mathscr{S}$ and vehicle $v \in \mathscr{V}$
$\overline{\Phi}_a$	Expected number of customers to come from period $t+1$ until period T originating from area $a \in \mathscr{A}$
Q	Vehicle capacity
r	Expected customers' profit
$\widetilde{s}_{as}$	$=1$, if the value function is approximated for state $x + 1_{as}$, 0 otherwise
τ_c	Service time per customer
τ_d	Time needed to travel one unit of distance
v_{as}	Attraction value for time slot $s \in \mathscr{S} \cup \{0\}$ of customers from area $a \in \mathscr{A}$ including the no-purchase alternative
w_{as}	Level of dissatisfaction if time slot $s \in \mathscr{S}$ is not offered to customers from area $a \in \mathscr{A}$

Decision variables

$\sigma^v_{as} \in \{0,1\}$	$=1$, if vehicle $v \in \mathscr{V}$ travels to ATC (a,s) for all areas $a \in \mathscr{A}$ and time slots $s \in \mathscr{S}$, 0 otherwise
$\delta^v \in \{0,1\}$	$=1$, if the current customer request is served by vehicle $v \in \mathscr{V}$, 0 otherwise
$\widetilde{D}^v_s \geqslant 0$	Expected travel distance in time slot $s \in \mathscr{S}$ for vehicle $v \in \mathscr{V}$
$\psi^v_{as} \geqslant 0$	Expected number of customers in ATC (a,s) for all areas $a \in \mathscr{A}$ and time slots $s \in \mathscr{S}$ which are served by vehicle $v \in \mathscr{V}$
$\varphi_{as} \geqslant 0$	Number of future customer requests from area $a \in \mathscr{A}$ that choose time slot $s \in \mathscr{S} \cup \{0\}$ s

References

Agatz, N., A. Campbell, M. Fleischmann, and M. Savelsbergh
2011. Time Slot Management in Attended Home Delivery. *Transportation Science*, 45(3):435–449.

Agatz, N., A. M. Campbell, M. Fleischmann, J. Van Nunen, and M. Savelsbergh
2013. Revenue management opportunities for Internet retailers. *Journal of Revenue and Pricing Management*, 12(2):128–138.

Asdemir, K., V. S. Jacob, and R. Krishnan
2009. Dynamic pricing of multiple home delivery options. *European Journal of Operational Research*, 196(1):246–257.

Campbell, A. M. and M. Savelsbergh
2004. Efficient Insertion Heuristics for Vehicle Routing and Scheduling Problems. *Transportation Science*, 38(3):369–378.

Campbell, A. M. and M. Savelsbergh
2005. Decision Support for Consumer Direct Grocery Initiatives. *Transportation Science*.

Campbell, A. M. and M. Savelsbergh
2006. Incentive Schemes for Attended Home Delivery Services. *Transportation Science*, 40(3):327–341.

Cleophas, C. and J. F. Ehmke
2014. When Are Deliveries Profitable?: Considering Order Value and Transport Capacity in Demand Fulfillment for Last-Mile Deliveries in Metropolitan Areas. *Business and Information Systems Engineering*, 6(3):153–163.

Ehmke, J. F. and A. M. Campbell
2014. Customer acceptance mechanisms for home deliveries in metropolitan areas. *European Journal of Operational Research*, 233(1):193–207.

Gallego, G., R. Ratliff, and S. Shebalov
2015. A General Attraction Model and Sales-Based Linear Program for Network Revenue Management Under Customer Choice. *Operations Research*, 63(1):212–232.

Hernandez, F., M. Gendreau, and J.-Y. Potvin
2014. Heuristics for Time Slot Management: A Periodic Vehicle Routing Problem View. Technical report.

Klein, R., J. Mackert, M. Neugebauer, and C. Steinhardt
2017a. A model-based approximation of opportunity cost for dynamic pricing in attended home delivery. *OR Spectrum*, Pp. 1–28.

Klein, R., M. Neugebauer, D. Ratkovitch, and C. Steinhardt
2017b. Differentiated Time Slot Pricing under Routing Considerations in Attended Home Delivery. *Transportation Science*, P. Forthcoming.

Koch, S. and R. Klein
2017. Time Window Budgets for Anticipation in Attended Home Delivery. *Ssrn*, (October):1–48.

Köhler, C., J. F. Ehmke, and A. M. Campbell
2018. Flexible Time Window Management for Attended Home Deliveries. (2).

Mackert, J.
2019. Choice-based dynamic time slot management in attended home delivery. *Computers & Industrial Engineering*, 129:333–345.

Sandoval, G.
2007. Webvan delivers its last word: Bankruptcy.

Yang, X. and A. K. Strauss
2017. An approximate dynamic programming approach to attended home delivery management. *European Journal of Operational Research*, 263(3):935–945.

Yang, X., A. K. Strauss, C. S. M. Currie, and R. Eglese
2016. Choice-Based Demand Management and Vehicle Routing in E-Fulfillment. *Transportation Science*, 50(2):473–488.